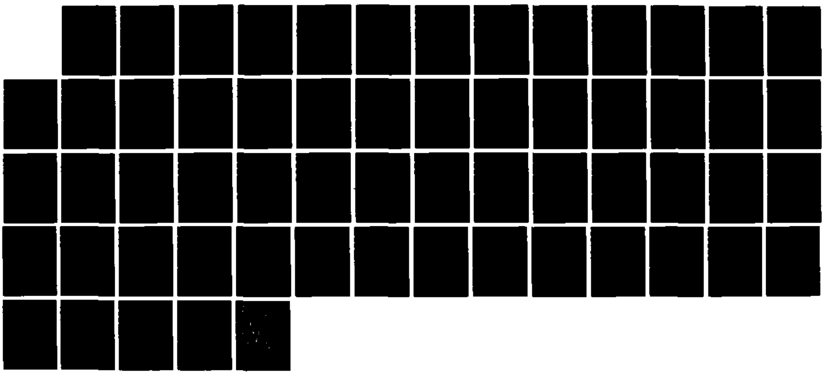
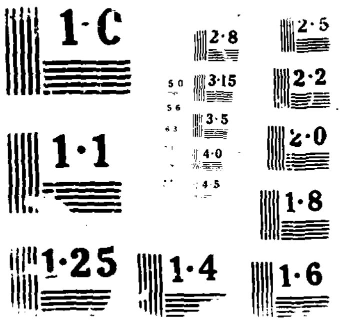
AD-A189 498 MAINTENANCE IN PROBABILISTIC KNOWLEDGE-BASED SYSTEMS 1/1
(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH
SCHOOL OF ENGINEERING T F REID DEC 87
UNCLASSIFIED AFIT/GOR/ENS/87D-16 F/G 12/5 NL




1.0
2.8
2.5
3.15
2.2
3.5
1.1
2.0
4.0
4.5
1.8
1.25
1.4
1.6

1

AFIT/GOR/ENS/86D-16

MAINTENANCE IN PROBABILISTIC
KNOWLEDGE-BASED SYSTEMS

THESIS

Thomas F. Reid
Captain, USAF

AFIT/GOR/ENS/86D-16

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY 2b DECLASSIFICATION/DOWNGRADING SCHEDULE	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
4 PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GOR/ENS/87D-16	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION School of Engineering 6b OFFICE SYMBOL *(If applicable)* AFIT/ENS	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS *(City, State, and ZIP Code)* Air Force Institute of Technology (AU) Wright-Patterson AFB, Ohio 45433-6583	7b. ADDRESS *(City, State, and ZIP Code)*
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b OFFICE SYMBOL *(If applicable)*	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS *(City, State, and ZIP Code)*	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11. TITLE *(Include Security Classification)*
MAINTENANCE IN PROBABILISTIC KNOWLEDGE-BASED SYSTEMS

12 PERSONAL AUTHOR(S)
Thomas F. Reid, B.S., Capt, USAF

13a. TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
MS Thesis	FROM ______ TO ______	1987 December	58

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18 SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	Artificial Intelligence, Expert Systems, Decision Analysis, Influence Diagram, Probability
12	09		

19 ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Thesis Chairman: Gregory S. Parnell, Lieutenant Colonel, USAF
Assistant Professor of Operational Sciences

Lynn Wolaver 24 Feb 88

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL Gregory S. Parnell, PhD	22b TELEPHONE *(Include Area Code)* (513) 255-2549 22c OFFICE SYMBOL AFIT/ENS

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. Recent developments using directed acyclical graphs (i.e., influence diagrams and Bayesian networks) for knowledge representation have lessened the problems of using probability in knowledge-based systems (KBS). Most current research involves the efficient propagation of new evidence, but little has been done concerning the maintenance of domain-specific knowledge, which includes the probabilistic information about the problem domain. By making use of conditional independencies represented in the graphs, however, probability assessments are required only for certain variables when the knowledge base is updated.

The purpose of this study was to investigate, for those variables which require probability assessments, ways to reduce the amount of new knowledge required from the expert when updating probabilistic information in a probabilistic knowledge-based system. Three special cases (ignored outcome, split outcome, and assumed constant outcome) were identified under which many of the original probabilities (those already in the knowledge-base) do not need to be reassessed when maintenance is required.

Although some reduction in the number of probability assessments can be achieved when the special cases apply, it appears other areas may be more productive in reducing the level of effort needed to maintain probabilistic KBS's. Topics recommended for future research include the development of efficient propagation techniques for multiply connected graphs, and investigation of methods to make the probability encoding process more efficient.

AFIT/GOR/ENS/86D-16

MAINTENANCE IN PROBABILISTIC KNOWLEDGE-BASED SYSTEMS

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Operations Research

Thomas F. Reid, B.S.

Captain, USAF

December 1987

Approved for public release; distribution unlimited

Preface

The desire to make the use of probability to represent uncertainty in knowledge-based systems more appealing was the primary motivation for this research effort. Credit for any success in achieving this goal must, however, be shared with the many fine people who helped me throughout this effort.

I am deeply indebted to my thesis advisor, Lt Col Gregory S. Parnell, and my primary reader, Maj Bruce W. Morlan. Throughout this research effort, they provided invaluable guidance, inspiration, and moral support. Our discussions were always invigorating and a constant source of new ideas. My other committee members, Dr. Frank Brown and Dr. Gary Lamont, greatly increased the quality of this thesis through their comments and insights. I also wish to thank Dr. John S. Breese, for his assistance in obtaining and using ALTERID, and Capt Jack Tomlinson and Lt Rick Miller of the Avionics Lab, for their help on the Symbolics.

I am grateful to all of my fellow classmates, from whom I learned a great deal; however, a few deserve special recognition. Many thanks to Capt Tom Burwell, for his efforts on the PERFORMA influence diagram tool, which was extremely useful in obtaining insights into the workings of influence diagrams. I am also greatly indebted to Captains Michael Ryan and David A. Drake for their moral support and encouragement throughout this research effort.

Finally, I owe the greatest debt to my wife, Deborah, for her understanding, support, and duties as "editor-in-chief" of my thesis.

Thomas F. Reid

A-1

Table of Contents

		Page
Preface		ii
List of Figures		v
List of Tables		vi
Abstract		vii
I.	Background	1
	Concerns About Probability in KBS	1
	Specific Problem	8
	Research Question	8
	Subsidiary Questions	8
	Scope	9
	Summary	10
II.	Current Status of Probabilistic Knowledge-Based Systems	11
	Influence Diagrams and Belief Networks	11
	Definitions and Notation	13
	Advantages of Influence Diagrams	15
	Propagation Using Influence Diagrams	19
	Summary	23
III.	Research Methodology and Special Case Development	24
	What Happens When the State of Information Changes?	25
	Taking the First Step	26
	Special Cases for Marginal and Conditional Probability Distributions	26
	Summary	40

IV. Conclusions and Recommendations 41

Conclusions 41
Areas for Future Research 43
Summary 45

Bibliography 46

Vita 48

List of Figures

Figure		Page
1	Influence Diagram for Metastatic Cancer Model	13
2	I.D. for Metastatic Cancer Model (Evidential Direction)	16
3	Ignored Outcome Versus General Case Data Requirements When Expanding A by k Outcomes	32
4	Split Outcome Versus General Case Data Requirements When Splitting an Outcome of A Into k Outcomes	35
5	Assumed Constant Outcome Versus General Case Data Requirements When Adding New Variable A	39

List of Tables

Table		Page
1	Why Not Probability in KBS?	2
2	Probabilities for Metastatic Cancer Model	14

AFIT/GOR/ENS/86D-16

Abstract

Recent developments using directed acyclical graphs (i.e., influence diagrams and Bayesian networks) for knowledge representation have lessened the problems of using probability in knowledge-based systems (KBS). Most current research involves the efficient propagation of new evidence, but little has been done concerning the maintenance of domain-specific knowledge, which includes the probabilistic information about the problem domain. By making use of conditional independencies represented in the graphs, however, probability assessments are required only for certain variables when the knowledge base is updated.

The purpose of this study was to investigate, for those variables which require probability assessments, ways to reduce the amount of new knowledge required from the expert when updating probabilistic information in a probabilistic knowledge-based system. Three special cases (ignored outcome, split outcome, and assumed constant outcome) were identified under which many of the original probabilities (those already in the knowledge-base) do not need to be reassessed when maintenance is required.

Although some reduction in the number of probability assessments can be achieved when the special cases apply, it appears other areas may be more productive in reducing the level of effort needed to maintain probabilistic KBS's. Topics recommended for future research include the development of efficient propagation techniques for multiply connected graphs, and investigation of methods to make the probability encoding process more efficient.

MAINTENANCE OF PROBABILISTIC KNOWLEDGE-BASED SYSTEMS

I. Background

Recent years have shown a growing use of artificial intelligence (AI) techniques, in particular that of knowledge-based (expert) systems (KBS), as an aid in the decision making process. One of the current areas of interest in AI is the representation of uncertainty because, as Lindley states, " . . . we want to study uncertainty . . . to be able to make decisions in the face of uncertainty" (6:130). Due to difficulties in the use of probabilities in knowledge-based systems, most of the AI systems in use today either ignore uncertainty in the problem domain or use non-probabilistic approaches for representing uncertainty. However, recent research has been somewhat successful in reducing, even eliminating, some of these difficulties (1; 7).

Concerns About Probability in KBS

Why has probability been largely ignored in most of the knowledge-based systems to date? Henrion (3:4), Pearl (7:242,252), and Rich (9:192-193) discuss reasons why Bayesian probabilities have not been used in most knowledge-based systems (Table 1). While Rich only lists reasons for not using probabilities, Henrion and Pearl present rebuttals to some of the arguments against the use of probabilities in knowledge-based systems. The arguments against using probability are addressed to varying degrees in the literature; some are shown not to be inherent problems in the use of probability, while others are still open questions.

Table 1. Why Not Probability in KBS?

	Reason	Henrion	Pearl	Rich
1.	Requires unrealistic independence assumptions	X		
2.	Can not handle second order uncertainty	X		
3.	Not how people do it	X	X	
4.	Computationally intractable	X	X	X
5.	Inference process hard to explain	X		
6.	Requires vast amounts of data (collection too hard)	X		X
7.	Method used for representation of uncertainty does not matter	X		X
8.	Difficult to modify knowledge base due to large number of complex interactions			X

1. Unrealistic Independence Assumptions. One critical distinction must be noted when reading the literature regarding the issue of independence assumptions: for probabilistic knowledge-based systems, a particular implementation may make independence assumptions to avoid computational complexity, but probability theory can readily handle dependencies among sources of evidence. So, while these arguments may be valid for some specific knowledge-based systems which use probability (e.g., Prospector), they do not hold for all such systems (e.g., ALTERID) (9:193; 1:164). Chapter II shows how knowledge-based systems based on influence diagrams can update belief based on dependent sources of evidence.

As Henrion shows (3:6-8), other methods of representing uncertainty (e.g., certainty factors or fuzzy set theory) make assumptions about dependence/independence of the sources of evidence. For example, the *and* and *or* fuzzy set operators assume the maximum possible correlation between evidential sources. Thus the view that fuzzy set

theory requires no assumptions about dependence (14:78,79) is misleading. There is merely no terminology, no concept, for such evidential dependencies in fuzzy set theory (3:7).

2. Inability to Represent Second Order Uncertainty. Another of the arguments against the use of probability in knowledge-based systems is that there is no mechanism for dealing with second order uncertainty, that is, uncertainty about the probabilities. An example would be an expert who was unwilling or unable to provide the knowledge engineer with a definite probability number: "Well, the chances of A happening lie somewhere around .45, but could range anywhere from .3 to .5." Henrion (3:5) states that using the mean value to estimate the probability ". . . is often sufficient . . . unless decisions about gathering new information are being contemplated," or that a range of probabilities may be specified to represent this second order uncertainty.

3. People Do Not Use Probability for Reasoning. The critique that "probability should not be used to represent uncertainty in knowledge-based systems because it does not reflect the way people reason" lies at the very heart of the disagreement between proponents of alternate uncertainty representations, such as fuzzy set theory, and proponents of Bayesian probability. Close examination of the many discussions dealing with the various representations for uncertainty (5; 6; 12; 14) reveals that the primary basis for disagreement over which representation is "best" is one of philosophy rather than one of method. The proponents of probability adopt a *normative* view of decision making, saying that it is better to represent uncertainty, not as people usually do, but rather as they should do if they desire to act in a consistent and logical manner. Proponents of the other representations subscribe to the *descriptive* approach, which promotes the view that uncertainty should be represented in a manner compatible to the

way people actually represent uncertainty. This difference of philosophy is clearly the primary issue in Zadeh's rebuttal (6:24) to Lindley's claim that ". . . the only satisfactory description of uncertainty is probability" (5:113). It is quite likely that this basic, philosophical issue is one of the stumbling blocks in the universal adoption of any one method for representing uncertainty in knowledge-based systems.

4. Computationally Intractable. One of the critical complaints about using probability in knowledge-based systems is the computational requirements. For example, the full joint distribution for n variables, each with two possible outcomes, contains 2^n probabilities. Reducing this exponential complexity has been the primary area of research in the use of probabilities in knowledge-based systems. Pearl has been especially active in this area, applying the use of directed, acyclical graphs (Bayesian nets) to the problem. Thus far, his results are limited to problems which can be represented in singly connected directed graphs, where there is no more than one (undirected) path between any two nodes (7:249).

Another representation which may somewhat alleviate this problem is the use of influence diagrams in implementing a probabilistic inference scheme in knowledge-based systems (1). Originally developed as a tool for decision analysis, influence diagrams have been used as an integral part of ALTERID, a tool for building knowledge-based systems which explores the ability to combine probabilistic and logical inference in the same system. In these systems, the individual evidences and hypotheses (conclusions) are represented as nodes in a directed graph. The arcs between the nodes represent dependencies between the probabilities for those nodes.

The principal difference between influence diagrams and Bayesian nets is the availability of decision and value nodes in the influence diagram. These nodes allow the use

of normative decision theory (i.e., maximum expected utility) in the decision process. Shachter (10:8-16) gives an algorithm which can be used for probabilistic inferencing. Chapter II briefly describes how influence diagrams and Shachter's algorithm may be used in probabilistic knowledge-based systems.

5. *Inference Process is Hard to Explain.* In order to effectively use knowledge-based systems in the decision making process, it is necessary to gain the user's confidence in the system. In many cases, this requires that the user understands the process by which the recommended action is obtained. Henrion gives arguments from Pearl and Spiegelhalter showing how this can be achieved using either 1) the logarithm of the likelihood ratio as an additive, relative measure of the impact new evidence has on a given conclusion; or 2) by stepping through the inference network (i.e., influence diagram or Bayesian network), where a "simple, intuitively meaningful" explanation can be given at each step (3:5). If, as Pearl argues, people reason based on "low-order marginal and conditional probabilities defined over small clusters of propositions," then in most cases the underlying inference network, obtained from the expert, is relatively sparse. Thus the explanation at each step involves relatively few propositions, making for a more intuitive description.

6. *Data Requirements.* Henrion attributes the mistaken view that vast amounts of data are required for the use of probabilities to a frequency-based interpretation of probability. This "problem" is not applicable to the Bayesian view of probability, where probability is a measure of an individual's degree of belief, not necessarily an absolute measure obtained by statistical sampling (3:5). Henrion also claims that the number of probabilities which must be encoded from the expert will not be inordinate if the infer-

ence structure reflects the way the expert thinks about the problem, a view shared by Shachter and Heckerman (11:56).

This, however, does not imply that it is a simple matter to encode probabilities—on the contrary, decision analysts (4:30-40) find this process to be as time consuming and as full of pitfalls as any part of the knowledge engineering process for rule-base systems. It is important to note that the difficulty does not lie in obtaining a number or even a set of numbers. The hard part is obtaining the correct numbers: those that form a coherent set, relative to the laws of probability. It is in this requirement for coherence that probability differs most drastically from other methods of representing uncertainty.

7. *Method Used to Represent Uncertainty Does Not Matter.* Of critical importance is the complaint that the method chosen for the representation of uncertainty, be it certainty factors, fuzzy set theory, or probability, really does not have much of an effect on the final action recommended by an expert system. Henrion briefly reviews two studies done on the MYCIN system (3:13-14). When the standard MYCIN certainty factor method for combining evidence was used, one-fourth of the cases indicated that certain evidence weakened the chance of a given hypothesis, when, in fact, that outcome was more likely, given the evidence! The other study examined granularity of the certainty factors, where changing from a continuous scale to a five-point scale (-1, -.5, 0, .5, 1) caused 9 out of 10 organisms to be misidentified. Although these studies do not fully answer the question of whether the choice of uncertainty representation makes a difference in the final result, they certainly raise questions about the common belief that it makes no difference. Henrion notes the lack of information in this area, and calls for further study.

8. Difficult to Modify the Knowledge Base. One difficulty of probabilistic knowledge-based systems that has received relatively little attention in the literature is that of maintaining the domain-specific knowledge base, which includes all of the probabilistic information. The use of probability to represent uncertainty does not mean that once such a knowledge-based system is built it will never need to be changed. As Waterman so eloquently states, "Expert systems . . . will make mistakes" (13:30). Just as rule-based systems must be maintained throughout their use, probabilistic knowledge-based systems are subject to changes in the knowledge base. This maintenance effort may be caused by changes in the underlying problem domain, an incorrect designation of the outcome space, or even an error in the initial encoding of the probabilities.

Heckerman and Horvitz (2:125) briefly address this concern. They argue that the steps necessary in maintaining the underlying inference structure (Bayesian network) include: 1) reassessment of the dependency structure by the expert, where changes in the arcs, node outcomes, or number of nodes are noted; and 2) reassessment of the probability distribution for each node whose incoming arcs changed. This holds for either adding or deleting a node. Little more than this brief discussion of probabilistic knowledge-based system maintenance is reported in the literature.

As indicated in this brief overview, research is relieving some of the concerns about using probability in knowledge-based systems. The work of Breese, Pearl, Shachter, and others is leading to more efficient methods of updating probabilities in knowledge-based systems through the use of Bayesian networks or influence diagrams. However, one key area that has not been adequately addressed is the maintenance of the underlying inference network in such knowledge bases.

Specific Problem

There has been little study of methods which can be used to minimize the impact of modifying the domain-specific knowledge base of a probabilistic KBS. This process, which involves obtaining further information from the expert(s) through knowledge engineering, can be very time-intensive. As more efficient methods for using probabilities in knowledge-based systems become available through the use of the new representation schemes, and as these systems are used in domains which may require more changes to the underlying knowledge base, reductions in the level of effort required to obtain new information become more critical.

Research Question

For knowledge-based systems which use probability to represent uncertainty, how can changes to probabilistic information in the domain-specific knowledge bases be managed to reduce the level of effort required in updating the system?

Subsidiary Questions

This research question can be partitioned into the following three subsidiary questions The information necessary to answer the research question is provided by answering all of these questions.

What changes can occur in probabilistic representations, and what are the corresponding changes to the knowledge representation of a probabilistic knowledge-based system?

Which types of changes may lend themselves to a reduction in the effort required to encode new probabilities, and how much reduction may be achieved?

For knowledge system tools which use influence diagrams to represent probabilistic knowledge, such as ALTERID, how can the knowledge system builder make use of the effort-saving cases identified in the second subsidiary question?

Scope

This investigation uses the following assumptions as a starting point in the research effort. These underlying assumptions are critical to this research effort.

1. The use of probabilities to represent uncertainty in knowledge systems is desirable—or, in some cases, necessary—in order to adequately model the uncertainties present in the problem domain and enable normative analysis. Additionally, these probabilities can be encoded from the expert's beliefs, as a part of the knowledge engineering process, but this research effort does not examine methods for encoding these probabilities. This probabilistic information is based on a specific *state of information*, that is, all of the knowledge which the expert used in determining the probabilistic information. This Bayesian approach to probabilities is consistent with much of the current research (1; 7; 3; 10).

2. Changes in the domain-specific knowledge reflect changes in the state of information available about the problem domain. For example, a test is developed, after the initial probabilities are encoded, which can provide additional evidence about some hypothesis of interest, or new information changes the expert's beliefs about the conditioning effects between the propositions in the knowledge system.

Using these assumptions as a basis, the research questions are directed at the issues concerning the difficulty of obtaining probabilities and the complexity in the database modification process. By reducing the quantity of information needed, a reduction in the effort required in modifying the knowledge base may be achieved.

Summary

The major difficulties of using probabilities in a KBS, as described in this chapter, motivated this research effort. Because the assessment of probabilities is a difficult and time-consuming process, it is desirable to reduce the number of assessments which are required during the maintenance of probabilistic knowledge-based systems. To define the framework for this discussion, Chapter II reviews the current research dealing with the use of probability and directed graphs (such as influence diagrams) to update beliefs in knowledge-based systems. Chapter III presents special cases under which fewer probability assessments *may be required during the maintenance effort*, while Chapter IV gives the conclusions reached during this thesis effort, as well as possible areas for further research.

II. Current Status of Probabilistic Knowledge-Based Systems

Before examining the maintenance of probabilistic knowledge-based systems, it is important to understand the process used to propagate the effects of evidence throughout such systems. Much of the current research in the use of probabilities to represent uncertainty in knowledge-based systems centers on two graphical representations: belief networks (also called Bayesian networks) and influence diagrams (1; 2; 7; 10). This chapter defines terms to assist in the discussion of these representations, and presents Shachter's algorithm, a flexible method for conducting probabilistic inference.

Influence Diagrams and Belief Networks

Both influence diagrams and belief networks contain information at two levels. On the most visible level, the network graph consists of nodes corresponding to each variable (or proposition) in the system, and arcs indicating dependencies between those nodes. On a lower level, more detailed information, such as probability distributions and a set of possible outcomes (possible values which the variable may take), are associated with each node. As is common in the literature (10:3), no distinction between a node and its associated variable is made throughout the remainder of this thesis.

While both influence diagrams and belief networks contain probabilistic (or chance) nodes, the influence diagram may also contain decision nodes, which maximize the expected value of their predecessors, and a value node, which represents a deterministic function of its predecessors (either chance or decision nodes). Influence diagrams form a super-set of belief networks. In fact, an influence diagram which contains no value or decision nodes (called a probabilistic influence diagram) is equivalent to a belief network (10:3).

Some propagation schemes for belief networks are limited to singly connected networks, where there is, at most, one undirected path between any two nodes (7:249). This limitation enables the use of *local propagation*, with its gains in efficiency and applicability to parallel processing, for a possible real-time inference capability (7:269). This capability is highly desirable, but many real-world problems can not obviously be represented by a singly connected network. One of the most promising approaches to overcome this limitation is the addition of auxiliary variables to convert a multiply connected network into a singly connected one (7:269-270). Unfortunately, it is not clear how these auxiliary variables should be added, what meaning these variables would have, or even if adding such nodes is always feasible (3:5). Another approach entails collapsing the multiply connected portion of the graph into a single variable, with outcomes corresponding to combinations of the collapsed variables' outcomes. However, this method is liable to exponential complexity in the number of collapsed variables.

For the remainder of this thesis, the discussion will center on the use of influence diagrams (vice belief networks) in knowledge-based systems. The reason for this focus is simple: influence diagrams may be used to represent any belief network, while only the probabilistic portion of an influence diagram can be represented in a belief network. The influence diagram propagation scheme can be used to solve belief diagrams for updated probabilities, so results which apply to the probabilistic portion of influence diagrams are equally applicable to belief networks. Also, most of the literature concerning belief networks is relevant to the use of influence diagrams.

Definitions and Notation

Before further examination of influence diagrams and their use in knowledge-based systems, a few helpful definitions and some notation for discussing influence diagrams is necessary. A more in-depth discussion can be found in Shachter (10:3-7).

The Metastatic Cancer Influence Diagram. The metastatic cancer influence diagram (depicted in Figure 1) is used to illustrate the definitions presented in this section, and to demonstrate other concepts throughout the remainder of this thesis.

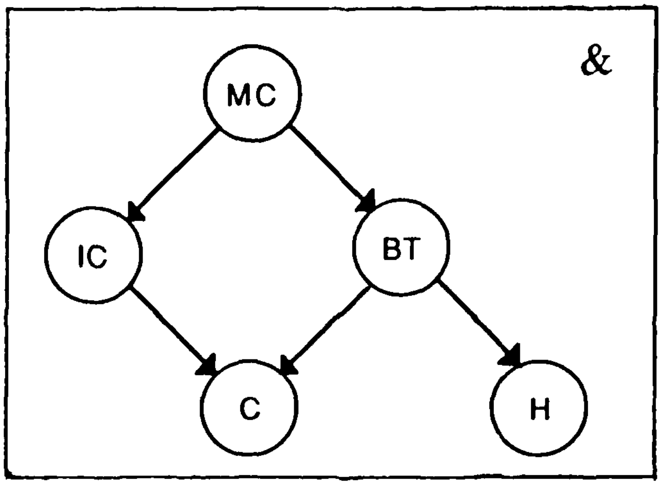


Figure 1. Influence Diagram for Metastatic Cancer Model

This model, also used by Pearl (8:248), is constructed in a causal direction. Based on the expert's current state of information (&), the presence of metastatic cancer (MC=mc) in a patient can possibly cause increased serum calcium (IC=ic), a brain tumor (BT=bt), or both. Either of these two can cause the patient to lapse into a coma (C=c). In addition, a brain tumor can be the cause of severe headaches (H=h) for the patient. The associated marginal and conditional probabilities are given in Table 2.

Table 2. Probabilities for Metastatic Cancer Model

P(MC | &):	P(mc | &) = 0.20	
P(IC | MC,&):	P(ic | mc,&) = 0.80	P(ic |¬mc,&) = 0.20
P(BT | MC,&):	P(bt | mc,&) = 0.20	P(bt |¬mc,&) = 0.05
P(C | IC,BT,&):	P(c | ic, bt,&) = 0.80	P(c | ic,¬bt,&) = 0.80
	P(c |¬ic, bt,&) = 0.80	P(c |¬ic,¬bt,&) = 0.05
P(H | BT,&):	P(h | bt,&) = 0.80	P(h |¬bt,&) = 0.60

This causal representation is just one of the 5! logically equivalent influence diagram representations for this problem (4:723). Each of these 5! representations result in the same joint distribution.

Definitions for Influence Diagrams. As indicated in Table 2, the probability distribution for the outcomes of a node (variable) is, in many cases, a conditional probability distribution. The conditioning for a node, I, is shown in the influence diagram by arcs into I from each of the *conditional predecessors*. The notation C(I) is used to represent the set of conditional predecessors of node I; that is, the set of all nodes with an arc going directly into node I. In the metastatic cancer example, the conditional predecessors for the coma node (C) are the increased serum calcium (IC) and brain tumor (BT) nodes. If C(I) is the empty set, then a *marginal distribution* is indicated, since I has no incoming arcs. The metastatic cancer node (MC) is an example of a node with a marginal distribution.

The set of *weak predecessors* of a node I, denoted by W(I), is defined as all nodes, J, for which there is a directed path from J to I. C(I) is a subset of W(I). For the metastatic cancer influence diagram headache node (H), C(H) is the set {BT}, and W(H) is {BT, MC}. Similarly, D(I), the *direct successors* of node I, is the set of nodes for which I is a conditional predecessor. For example, D(MC) is the set {IC, BT}.

The final concept that needs to be introduced is that of *node ordering*. A list of nodes is ordered if "none of the weak predecessors of a node follow the node in the list" (10:6). Any list of nodes can be ordered by placing a node with no conditional predecessors on the ordered list, then adding other nodes, one at a time, whose conditional predecessors, if any, are already on the ordered list (10:6-7). There may be more than one ordered list for any given influence diagram.

Advantages of Influence Diagrams

Heckerman and Horvitz (2) discuss the advantages of belief networks over rule-based representations. In particular, they show that such graphical representations are more natural and efficient in representing dependencies within the problem domain. There is, as they point out, an increased cost for representing more complex dependencies in any knowledge-based system: more probabilities must be encoded, and computational costs are greater. They claim, however, that rule-based representations are at least as costly, and usually more costly, than belief networks (and hence influence diagrams) in representing these dependencies (2:125).

Another advantage of influence diagrams is the ability to represent domain knowledge in a causal direction. More recent literature (11; 7) indicates that people find it easier to describe influences in a *causal direction*, that is, based on hypotheses which cause evidences. This is directly opposite of the direction this information is used in many rule-based systems, which require the probability of the hypothesis given the evidences (11:55). Recall the metastatic cancer influence diagram, which was constructed in the causal direction. The influence diagram depicted in Figure 2, which follows the evidential direction required for a rule-based system, is logically equivalent to the causal graph shown in Figure 1.

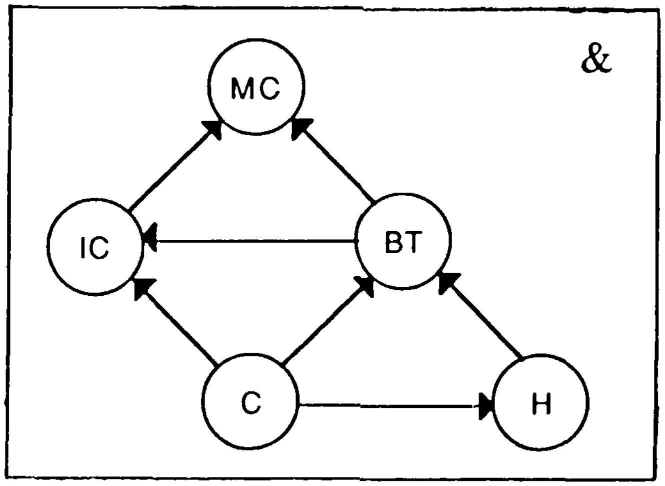


Figure 2. I.D. for Metastatic Cancer Model (Evidential Direction)

A typical rule base for this model, but one which does not capture the dependencies indicated in the influence diagram in Figure 2, might resemble the following:

IF IC=increased_calcium AND BT=brain_tumor
THEN MC=metastatic_cancer (.#)
IF IC=increased_calcium THEN MC=metastatic_cancer (.#)
IF BT=brain_tumor THEN MC=metastatic_cancer (.#)
IF C=coma AND H=severe_headaches THEN BT=brain_tumor (.#)
IF C=coma THEN IC=increased_calcium (.#)
IF C=coma THEN BT=brain_tumor (.#)
IF H=severe_headaches THEN BT=brain_tumor (.#)

These rules, and the associated degrees of belief (.#), would be acquired as part of the knowledge engineering process. Since all the dependencies are not represented by this rule-based representation, an ad hoc method for combining degrees of belief (i.e. MYCIN's certainty factor method) may be used, and the rules may need to be revised to achieve better results.

In order to accurately represent the dependencies shown in Figure 2, a larger rule base is required:

IF IC=increased_calcium AND BT=brain_tumor
THEN MC=metastatic_cancer (.##)
IF IC=increased_calcium AND BT=not_brain_tumor
THEN MC=metastatic_cancer (.##)
IF IC=not_increased_calcium AND BT=brain_tumor
THEN MC=metastatic_cancer (.##)
IF IC=not_increased_calcium AND BT=not_brain_tumor
THEN MC=metastatic_cancer (.##)
IF C=coma AND BT=brain_tumor
THEN IC=increased_serum_calcium (.##)
IF C=coma AND BT=not_brain_tumor
THEN IC=increased_serum_calcium (.##)
IF C=not_coma AND BT=brain_tumor
THEN IC=increased_serum_calcium (.##)
IF C=not_coma AND BT=not_brain_tumor
THEN IC=increased_serum_calcium (.##)
IF C=coma AND H=severe_headaches THEN BT=brain_tumor (.##)
IF C=coma AND H=not_severe_headaches THEN BT=brain_tumor (.##)
IF C=not_coma AND H=severe_headaches THEN BT=brain_tumor (.##)
IF C=not_coma AND H=not_severe_headaches
THEN BT=brain_tumor (.##)
IF C=coma THEN H=severe_headaches (.##)
IF C=not_coma THEN H=severe_headaches (.##)

These fourteen rules represent the dependencies shown in the influence diagram, and the degree of belief (.##) for each rule would be acquired as part of the knowledge engineering process. This compares with the required assessment of only eleven probabilities for the influence diagram representation, as shown in Table 2. More complex dependencies may cause the rule-based method to degenerate to a very basic look-up table, where no chaining would be possible since each combination of all evidences and hypotheses is represented by a corresponding rule (2:123). Thus, only one rule could possibly match.

One of the most important, and perhaps the most overlooked, advantages of the influence diagram is referred to by Heckerman and Horvitz (2:125) as a local notion of modularity. In actuality, this is just the *conditional independence* shown by the lack of an arc from one node into another in an influence diagram. Conditional independence of

two nodes exists when, given the values of the conditional predecessors of one node, the probability distribution of that node is independent of the value of the other node. As a consequence of the conditional independence represented in an influence diagram, the probability distribution of any node, given the values of its conditional predecessors, is independent of the other weak predecessors of that node. This can be written as

$$P[I=i \mid W(I),\&] = P[I=i \mid C(I),\&] \tag{1}$$

As Shachter (10:7), Heckerman, and Horvitz (2:125) state, the marginal and conditional probabilities contained in an influence diagram contain all the information necessary to construct the full joint distribution. If an influence diagram consists of the chance nodes I_1, I_2, ... , and I_m, then the joint distribution is simply the product of all the marginal and conditional distributions in the graph:

$$P[I_1=i_1, I_2=i_2, \ldots, I_m=i_m \mid \&] = \prod_{j=1}^{m} P[I_j=i_j \mid C(I_j),\&] \tag{2}$$

As Heckerman and Horvitz indicate, this property can greatly reduce the data collection associated with modifying the influence diagram. There is no need to reassess all of the probabilities in an influence diagram when a change occurs, since only the direct predecessors of a node influence its probability distribution. Only nodes which have their incoming arc(s) modified need to have their distributions reassessed. This particular property is critical to the development of the special cases in Chapter III.

To see the value of this property more clearly, consider the following change to the metastatic cancer example. The experts have determined that the severity of a patient's headaches may be caused by increased serum calcium as well as a brain tumor, so an arc is added from the increased serum calcium node (IC) to the severe headache node (H). The new joint distribution can be obtained either by 1) assessing the joint

distribution directly; 2) assessing all of the conditional and marginal distributions for the new influence diagram; or 3) assessing only the new distribution for H. To reassess the existing probabilities in the joint distribution directly would require 2^5-1 probability assessments. Only thirteen assessments are needed to define the marginal and conditional distributions for the new influence diagram, so some savings are realized by choosing to represent the joint as the product of marginal and conditional distributions. Even more savings are obtained because of local modularity, as only four assessments are required to define H's new conditional distribution.

Propagation Using Influence Diagrams

One of the main attractions of influence diagrams lies in the flexibility provided through the use of a few simple operations. Nowhere is this flexibility more evident than in Shachter's algorithm for solving any well-formed influence diagram (one containing no directed cycles). His algorithm either provides the solution of a general probabilistic inference problem (i.e., the conditional probability distribution of a set of hypotheses, given some set of evidences), or determines what data are needed if some of the probabilities or outcomes for the influence diagram have not been specified (10:8).

Before examining this algorithm, a short overview of some of the operations for manipulating influence diagrams is in order. Only the three primative operations required in order to perform probabilistic inference using influence diagrams (10:10-16) will be covered here; readers are referred to (10) and (4) for other operations. None of these three operations change the basic, underlying (conditional) probability distribution for the hypotheses given the evidences (10:10).

Barren Node Elimination. A node which has no successors and is neither a hypothesis nor an evidence is called a *barren node*. As Shachter says,

> Since we are not trying to estimate [its probability distribution], we do not observe its value, and no other variables are conditioned by it, it is irrelevant to the inference problem we are solving. Therefore, we could eliminate the barren node...without affecting the solution. [10:10]

Two important clarifications should be made. First, no node is inherently barren in an influence diagram. The definition of being barren is strictly related to a given inference problem. Take, for instance, the metastatic cancer example presented earlier. If the hypothesis is "Metastatic cancer is present," or, equivalently, that MC=mc, and the observable evidence is the presence or absence of severe headaches, then the coma node (C) is barren, as is the increased serum calcium node (IC) once C is removed. On the other hand, if the observable evidence is the patient's being in a comatose state or not, the only barren node is the severe headache node (H). If both C and H are observable, as would normally be the case, no node is barren without transforming the diagram further.

The other clarification relates to the state of information. While it could be argued that the state of information upon which the influence diagram is based has changed once a barren node is removed, this can be viewed more as a *coarsening* of data. That is, the probabilities of interest have not changed, but some information has been lost from the influence diagram. It would be quite impossible, given only what is contained in the reduced influence diagram, to reconstruct the original nodes, dependencies, and probabilities that were removed.

Arc Reversal. Perhaps the most important operation in the solution of a probabilistic inference problem is arc reversal. Arc reversals are used for two primary purposes in Shachter's algorithm. First, arc reversals are used to transform the influence diagram, making nodes which are neither hypotheses nor evidences barren, so they can be removed. The second purpose for using arc reversal is to obtain the conditional

distributions of interest. The only requirement for reversing an arc is that there be no other directed path between the nodes at either end of the arc.

When an arc from node I to node J is reversed, each node inherits the conditional predecessors of the other. Then, through summing and application of Bayes' law, the conditioning between I and J is reversed (10:13):

$$P[J=j \mid C_{new}(J),\&] = \sum_{\substack{outcomes \\ of\ I}} P[J=j \mid C_{old}(J),\&] \; P[I=i \mid C_{old}(I),\&] \tag{3}$$

$$P[I=i \mid C_{new}(I),\&] = \frac{P[J=j \mid C_{old}(J),\&] \; P[I=i \mid C_{old}(I),\&]}{P[J=j \mid C_{new}(J),\&]} \tag{4}$$

where $C_{new}(J)$ is the combined conditional predecessors of I and J, excluding the node I, and $C_{new}(I)$ is the new conditional predecessors of J and the node J itself.

Deterministic Node Propagation. To gain computational efficiencies, Shachter uses the notion of a *deterministic node,* which is a chance node with a degenerate probability distribution (10:4). Shachter proves that, for any influence diagram which is fully specified (has complete graphical, outcome, and probability distribution information) and contains an arc from a probabilistic node, I, to another node, J, it is possible to remove the conditioning of I relative to J (10:11). This is done simply by adding arcs from the conditional predecessors of I to node J, then removing the arc from I to J. This is possible since the outcomes of I are determined exactly by the outcomes of I's conditional predecessors. Additionally, if J is deterministic, it will remain deterministic.

Node Reduction. Removing nodes from the influence diagram without changing the underlying probability distribution for the hypotheses given the evidences, called node reduction, is a combination of the three simple operations above (10:14). Reducing a

deterministic node is extremely simple: merely propagate the deterministic node into each of its successors, then remove the (now barren) deterministic node.

Reducing a probabilistic node is only slightly more complicated. First, an ordered list of the direct successors of the node must be obtained. The goal node must be placed last in the ordered list if it is a direct successor of the node being reduced. Arcs are reversed in order, then the (now barren) node is removed.

Shachter's Algorithm. Although Shachter's algorithm may be used to solve any general probabilistic inference problem given a fully specified influence diagram, and is the most flexible of the available probabilistic inferencing schemes, it is very simple to execute. To find the solution to any such inference problem, even those with many variables of interest (i.e., multiple hypotheses) a new deterministic variable, conditioned on the variable(s) of interest, is added to the diagram as the goal node. The goal node has no successors, and will gain none throughout the solution process. After adding the goal node, all that remains is the reduction of all non-evidence nodes from the diagram (10:15). Although no particular order is required for the solution of the inference problem, always reducing barren nodes and conditional predecessors of the goal node allows the solution to be obtained with the smallest possible amount of information specified in the diagram (10:16). This would be highly desirable in situations where all of the probabilistic information was not known. A probabilistic KBS could then use Shachter's algorithm to request the minimum amount of information necessary to solve the specified probabilistic inference problem.

Use of Influence Diagrams in KBS. Up to now, this chapter has covered the mechanics of performing probabilistic inference in a general sense. The construction of the influence diagram from probabilistic nodes and arcs, desirable properties associated

with influence diagrams, the transformations required for probabilistic inferencing, and a simple, powerful algorithm for the solution of probabilistic inference problems have all been discussed. But just how can all of this be used in knowledge-based systems?

In many KBS in use today, actions are taken (or recommended) based on a degree of belief in a proposition, or a set of propositions. That degree of belief is altered through knowing the outcomes of other, observable propositions. Using an influence diagram, the entire conditional distribution for the hypotheses, given the evidences, can be computed in a simple manner. This distribution can then be used as the basis for the KBS' recommended action, or further normative analysis.

Summary

This chapter laid the basic foundation for understanding the use of influence diagrams in reasoning with uncertainty in knowledge-based systems. Influence diagrams provide an efficient means to represent dependencies among variables, and contain all the information needed to construct the joint distribution. Shachter's algorithm provides a simple, flexible way to perform probabilistic inferencing with influence diagrams. When an influence diagram must be modified, the local modularity property reduces the number of probabilities that must be assessed. The next chapter examines situations where further reductions may be possible.

III. Research Methodology and Special Case Development

Just as with rule-base systems, the general domain knowledge of probabilistic knowledge-based systems is not static. The underlying (domain-specific) knowledge in the KBS must be changed to reflect the changed state of information when new tests for existing hypotheses are developed, new hypotheses are formed, or a more thorough understanding of the problem domain is gained. Heckerman and Horvitz (2:125-126) briefly discuss the process of adding a node to a Bayesian belief network, but their discussion holds equally well for influence diagrams. When determined (by whatever means) that the model represented in the influence diagram is no longer adequate, the first task for the knowledge engineer and the expert is the reassessment of the nodes and their dependencies represented in the influence diagram. Nodes may be added or deleted, outcome spaces for individual variables may increase or decrease, arcs may be added or deleted, or the probability distributions for a variable's outcomes may be changed, any combination of which indicates a change in the state of information upon which the graph is based. Because of the local modularity property, after the expert reassesses the basic underlying dependency structure, the only probability distributions that must be re-encoded are those associated with nodes that have had some change made to their outcome space (gaining, losing, or changing outcomes) or incoming arcs (gaining or losing an incoming arc, or having the outcome space of a conditioning variable modified) (2:125).

In systems where the underlying dependency structure changes infrequently, and there is a requirement for real-time propagation of the effects of evidences, the time and effort required to encode the new distributions may be relatively insignificant. However,

as these systems are applied to problem domains which are highly dynamic, meaning the underlying dependencies and probabilities often change, a significantly larger portion of time will be spent encoding probabilities. An example of such a dynamic system might be a KBS which interprets intelligence data and attempts to identify specific enemy tactics. As the enemy develops new tactics, and analysts identify discriminators for indicating when these tactics are being used, new nodes and dependencies are added to the KBS. This chapter examines ways in which the underlying probabilistic information can change and possible means to reduce the level of effort required in the encoding process, relative to these changes.

What Happens When the State of Information Changes?

As indicated earlier, when the underlying state of information changes, the dependency structure for the influence diagram must be reassessed, and those nodes which experience a change in their incoming arcs or outcome space must be reassessed. The more nodes that experience such changes, the more information that must be encoded by the expert. At the very least, probabilistic information must be encoded for the new outcomes and variables. Also, any data invalidated by the change in the state of information must be reassessed, even if the dependency structure did not change. However, all is not necessarily lost. There may be some circumstances under which all, or nearly all, of the original probabilistic information is still valid under the new state of information. Some of these circumstances are identified in the following sections as special cases which may apply for some state of information changes for an influence diagram.

Taking the First Step

Since the joint distribution is fully represented by the information in an influence diagram, the initial focus of the research was on keeping the joint distribution from changing a great deal. Specifically, special cases were developed for: 1) changing the number of outcomes for individual nodes; and 2) changing the number of variables in the joint distribution. As these special cases were being developed, it became increasingly clear that this focus was not the best for purposes of this research effort. Examining the special cases for the joint distribution revealed that more than one type of change in the marginal and conditional probabilities could bring about the same change to the joint distribution. This multiplicity of causes obscured the conditions which might lead to the special case's being applicable. A more fundamental view, based on how the probabilistic information would be gathered in a probabilistic KBS, was adopted.

Special Cases for Marginal and Conditional Probability Distributions

As indicated by Pearl, Shachter, and others, information from experts is more easily gathered in the form of marginal and conditional distributions (3:5; 7:246; 11:55). Since information is primarily collected in this manner, it makes much more sense to examine possible effort-saving special cases from this perspective. The primary objective is to keep as many of the original probabilities as possible relevant under the new state of information.

Special cases based on the marginal and conditional distributions can be readily grouped into those applicable when 1) the outcome space for a variable changes in size; 2) a variable is added or removed from the influence diagram, and 3) an arc between two nodes is added or removed, changing the conditioning information in the diagram.

The only other change which indicates a new state of information is when underlying probabilities change. No special cases were found to reduce the number of assessments required in response to this type of change.

For each special case, we examine, separately, the effects on the node being changed (either a node experiencing a change in its outcome space, or a new, added node) and on nodes whose incoming arcs are somehow modified (either by a change in the outcome space of a conditional predecessor, or by the addition or loss of conditional predecessors). Application of a special case to one of these nodes neither ensures nor prohibits its application to another node. For each special case, the effects on the changed node and on its successors will be discussed. Since exponential growth can occur when changes are made to the influence diagram, these special cases were developed primarily with an expansion of the outcome space or number of variables in mind.

Changes in the Outcome Space. When the change to a new state of information results in a change in a node's outcome space, the probability distribution for that node must be reassessed. The distribution of any other nodes which were previously, or are now, conditioned on the changed node must also be reassessed. Two special cases, the "ignored outcome" and the "split outcome", may reduce the number of assessments required.

Ignored Outcome Special Case. Interest in the first special case was motivated by the following question: if a new, or previously "forgotten" outcome was added to a node (indicating a new state of information), under what conditions would the original probabilistic information be of use, and just how could it be used? For this case, the

original outcome space for the changed variable would be mutually exclusive but not collectively exhaustive.

A simple example clarifies the discussion of this case. Let A be a node with m outcomes under the original state of information, $\&$. Now the expert perceives a previously ignored outcome, a_{m+1}. This new knowledge (that outcome a_{m+1} exists) indicates a change in the state of information, and the diagram must be reassessed relative to this new state of information, $\&'$. If the expert determines that the old probability distribution for A, given the conditional predecessors $C(A)$, is

$$P[A=a_i \mid C(A),\&] = P[A=a_i \mid C(A),A \neq a_{m+1},\&'] \tag{4}$$

then the new probabilities for the original outcomes of A are given by

$$P'[A=a_i \mid C_j(A),\&'] = \lambda_j \; P[A=a_i \mid C_j(A),\&] \quad \begin{array}{l} i = 1,\ldots,m \\ j = 1,\ldots,\prod_{X \in C(A)} \|X\| \end{array} \tag{6}$$

where $\|X\|$ denotes the number of outcomes for variable X. λ_j is just a scaling factor for the probability distribution of A given the old state of information and a specific combination, indexed by the subscript j, of the outcomes for variables in $C(A)$, and is given by

$$\lambda_j = 1-P[A=a_{m+1} \mid C_j(A),\&'] \tag{7}$$

To see more clearly the use of λ_j, let A (with outcomes a_1 and a_2) have one conditional predecessor, X (with outcomes x_1 and x_2). Let the conditional probabilities of A be:

$$\begin{array}{ll} P[A=a_1 \mid X=x_1,\&] = \alpha_1 & P[A=a_2 \mid X=x_1,\&] = \alpha_2 = 1-\alpha_1 \\ P[A=a_1 \mid X=x_2,\&] = \beta_1 & P[A=a_2 \mid X=x_2,\&] = \beta_2 = 1-\beta_1 \end{array}$$

If, under a new state of information ($\&'$), A has a new outcome (a_3), and the ignored outcome special case is applicable, then the conditional probabilities can be found as follows. First, the conditional probabilities of the new outcome, given the outcomes of X, must be encoded from the expert. Say they are determined to be:

$$P[A=a_3 \mid X=x_1,\&'] = \alpha_3$$
$$P[A=a_3 \mid X=x_2,\&'] = \beta_3$$

After these probabilities have been encoded, the λ_j's are given by

$$\lambda_1 = 1-\alpha_3$$
$$\lambda_2 = 1-\beta_3$$

Then the conditional probabilities for the original two outcomes of A are given by

$$P[A=a_1 \mid X=x_1,\&'] = \lambda_1\alpha_1 \qquad P[A=a_2 \mid X=x_1,\&'] = \lambda_1\alpha_2$$
$$P[A=a_1 \mid X=x_2,\&'] = \lambda_2\beta_1 \qquad P[A=a_2 \mid X=x_2,\&'] = \lambda_2\beta_2$$

When considering the addition of k new outcomes (instead of just one), the primary difference deals with the calculation of the λ_j. For each possible combination of outcomes for $C(A)$, λ_j is given by

$$\lambda_j = 1 - \sum_{i=m+1}^{m+k} P[A=a_i \mid C_j(A),\&'] \tag{8}$$

This means, for the expanded variable A, only the marginal or conditional probabilities for the new outcomes must be encoded. Once these are obtained, a λ_j for each combination of outcomes of $C(A)$ can be computed directly, and the probabilities under $\&'$ for the original outcomes are given by Eq (6).

The reduction in the required number of encodings depends on the number of old (m) and new (k) outcomes for A, the number of conditional predecessors for A ($\|C(A)\|$), and the number of outcomes for each predecessor. For comparative pur-

poses, suppose that each conditional predecessor of A has n outcomes[1]. Then the number of encodings needed to determine A's distribution in the general case is $(m+k-1)\times n^{\|C(A)\|}$, since probabilities for all but one of the A's $m+k$ outcomes are needed for each combination of the outcomes in $C(A)$. The probability of the remaining outcome of A is determined by the property that these probabilities must sum to one. Similarly, the number of required probability assessments for the ignored outcome special case is just $k\times n^{\|C(A)\|}$, because only the probabilities for the new outcomes of A are needed.

A similar reduction can be found in the number of probability assessments for direct successors of nodes with increased outcome spaces. The applicability of the special case must be assessed for each direct successor node individually. Referring to the simple example given above, suppose B was a direct successor of A. When the probability distribution for B, under the new state of information $\&'$, is given by

$$P[B=b_j \mid A=a_i, C(B), \&'] = P[B=b_j \mid A=a_i, C(B), \&] \quad i=1, \ldots, m \tag{9}$$

where a_i is in the set of original outcomes for A, then the only conditional distributions which must be assessed for B are those which are conditioned on $a_{m+1}, \ldots, a_{m+k}$, the new outcomes of A. This means the original distributions for B, given the old outcome space for A, are still valid under the new state of information. For example, say the old outcome space for A was {sunny, rainy}, and B was {get_wet, not_get_wet}. If the outcome space for A is expanded to {sunny, rainy, snowy}, the expert must determine whether or not the old distribution for "getting wet" or not, given that it is sunny (or

[1]This supposition is only made for notational convenience. All results remain valid when the number of outcomes is allowed to vary for each conditional predecessor, but the number of combinations of those outcomes is calculated differently.

rainy), is still valid. Only those conditional distributions, associated with specific outcomes of A, which the expert finds not to be valid must be reassessed. At the very least, the conditional distributions for B, given the new outcomes of A, must be assessed.

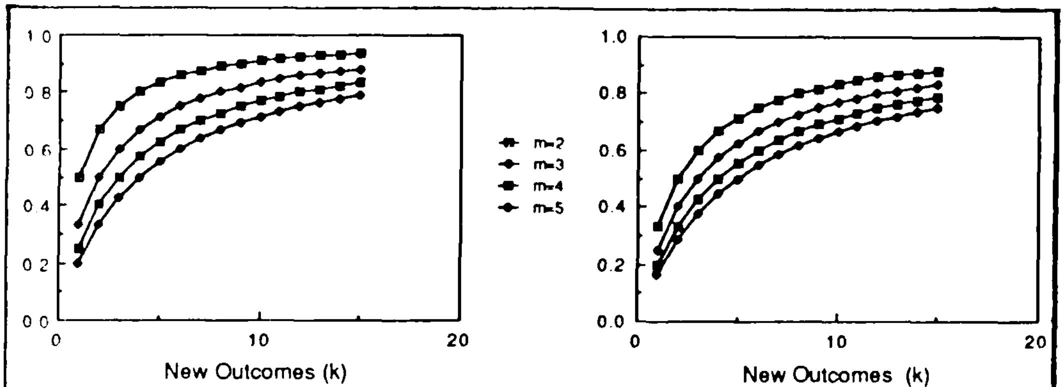
The number of probability assessments required to determine B's distribution depends on the m, k, the number of outcomes for B (p), and the number of outcomes for each of the variables in $C(B)\backslash A$[2]. For the general case, $(m+k)\times(p-1)\times n^{\|C(B)\backslash A\|}$ probability assessments are needed. This is reduced to $k\times(p-1)\times n^{\|C(B)\backslash A\|}$ when this special case applies.

To see these reductions more clearly, consider the metastatic cancer example (Figure 1) introduced in Chapter II. If the expert determines there are k additional outcomes of the BT node, then in general the number of probabilities which must be encoded to obtain the new distribution for BT is $(2+k-1)\times2^1$, or $2k+2$. Note that this is exponential in the number of conditional predecessors for BT (one, in this case). When this special case applies, this number is reduced to $2k$ since the two probabilities encoded under the original state of information do not need to be reassessed. The size of this reduction is equal to the amount of data already encoded under the original state of information. Once the new probabilities are encoded, the appropriate λ_j's can be computed using Eq (8). Similarly, for the C node, the number of assessments required to determine its distribution is, in general, $(2+k)\times(2-1)\times2$, or (2^2+2k). If the original conditional distributions for node C (which were based on the original set of outcomes for BT) are still valid, then only the probabilities for C which are conditioned

[2] $C(B)\backslash A$ denotes the set of all conditional predecessors of B, excluding A.

on the new outcomes of BT must be encoded, reducing the number of required assessments to $2k$. If this special case is applicable to node H as well, then the number of assessments needed to determine its distribution is reduced from $(2+k)\times(2-1)$ for the general case, to $(2-1)\times k$ for the ignored outcome special case.

Figures 3.a and 3.b show a graphical comparison of the relative number of assessments required in the general case and the number required in the ignored outcome case. Although these graphs show effects for up to 15 added outcomes, realistically only five or fewer outcomes would be added.



a. Relative Effectiveness of Ignored Outcome Special Case for Node A

b. Relative Effectiveness of Ignored Outcome Special Case for Node B

Figure 3. Ignored Outcome Versus General Case Data Requirements When Expanding A by k Outcomes

For each indicated value of m, the curves show the ratio of the number of assessments required for the special case to the number required in the general case. This ratio is $\frac{k}{m+k-1}$ for Figure 3.a, and $\frac{k}{m+k}$ for Figure 3.b. If A originally has two outcomes, then gains another due to a change in the state of information, Figure 3.a shows that the special case requires 1/2 as many probability assessments as the general case to

determine the distribution for A, and Figure 3.b shows that only 1/3 as many are needed to determine the distribution for B.

Notice that as the number of new outcomes increases, the relative effectiveness of the special case decreases: as $k \rightarrow \infty$ the special case requires nearly as many probability assessments as the general case for both nodes A and B. Conversely, as m increases, the relative effectiveness of the special case increases. Both of these reflect that the effectiveness of the special case depends on the amount of growth relative to the amount of data for the given distribution in the original influence diagram.

Split Outcome Special Case. A similar special case exists for situations where an outcome of a variable, say A, is split into two or more distinct outcomes. In effect, the original outcome, say $A = a_s$, was actually many outcomes: $a_{s1}, a_{s2}, \ldots, a_{sk}$. Unless the expert determines otherwise, the probabilities for the unchanged outcomes are still valid under the new state of information. That is,

$$P[A = a_i \mid C(A), \&'] = P[A = a_i \mid C(A), \&] \quad i \neq s \tag{10}$$

for a_i in the set of unchanged outcomes of A. The conditional probabilities for the new outcomes can then be assessed directly, where

$$\sum_{i=1}^{k} P[A = a_{si} \mid C(A), \&'] = P[A = a_s \mid C(A), \&] \tag{11}$$

for each combination of outcomes of the conditional predecessors of A. Alternatively, these probabilities may be gathered as fractions of $P[A = a_s \mid C(A), \&]$. The knowledge engineer would need to know, for each combination of $C(A)$, the probability that $A = a_{si}$ given that A is one of the new outcomes.

For either method of assessment, the sum of the new conditional probabilities is known, thus reducing by one the required number of assessments for each combination

of $C(A)$. This (minute) savings is not generally applicable for the ignored outcome special case, since there is no *a priori* way to determine the sum of the λ_j's.

Just as for the ignored outcome case, direct successors of a "split outcome" node do not have to have probabilities reassessed which are dependent on the unchanged outcomes of A. Only probabilities conditioned on the new a_{si} outcomes need be assessed. Thus the new distribution is given by

$$P[A=a_i \mid C(A),\&'] = P[A=a_i \mid C(A),\&] \quad i \neq s \tag{12}$$

for the unchanged outcomes of A, and

$$P'[A=a_{si} \mid C_j(A),\&'] = \lambda_{ij}\ P[A=a_s \mid C_j(A),\&] \quad \begin{array}{l} i = 1, \ldots, k \\ j = 1, \ldots, \prod_{X \in C(A)} \|X\| \end{array} \tag{13}$$

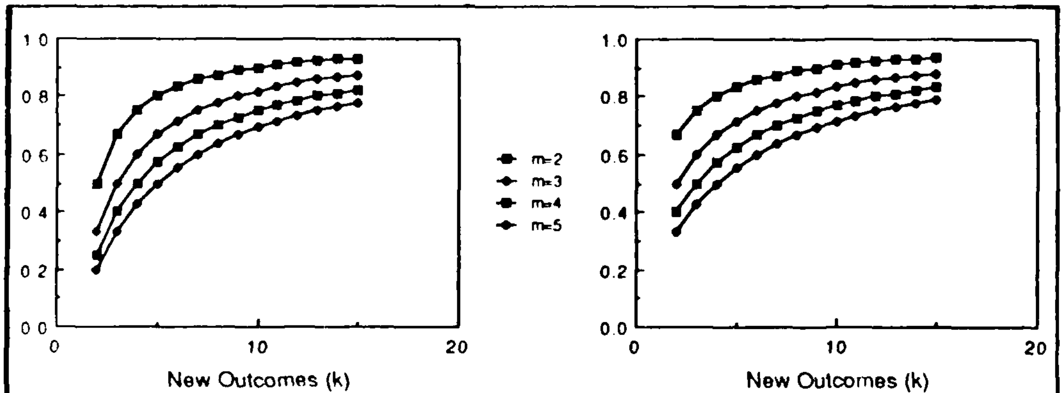
where $\sum_{i=1}^{k}\lambda_{ij}$ equals one for each combination $C_j(A)$ of A's conditional predecessors.

Since A has $m+k-1$ outcomes under the new state of information, the number of assessments required to form A's distribution (for the general case) is $(m+k-2)\times n^{\|C(A)\|}$. When the split outcome case applies, this is reduced to $(k-1)\times n^{\|C(A)\|}$ since the sum of the probabilities for the k new outcomes of A is known from Eq (12).

Just as for the ignored outcome special case, any direct successor (B) of A may not need all of its probabilities reassessed. If the the expert determines that the original conditional probabilities for B, given outcomes $a_1, \ldots, a_m$ of A, are still valid, then only the probabilities concerning A's new outcomes must be assessed. If B has p outcomes, and each variable in $C(B)\backslash A$ has n outcomes, then for the general case

$(m+k-1)\times(p-1)\times n^{\|C(B)\setminus A\|}$ probability assessments are required. If the split outcome case applies, the number of assessments is reduced to $k\times(p-1)\times n^{\|C(B)\setminus A\|}$.

The graphs in Figures 4.a and 4.b show the effectiveness of the split outcome special case relative to the general case. Similar to the graphs in Figure 3, these graphs show the ratio of the required number of assessments for the split outcome special case to those for the general case: $\frac{k-1}{m+k-2}$ for Figure 4.a, and $\frac{k}{m+k-1}$ for Figure 4.b. Again, notice the same type of effect from the relative size of the increase in A: as k becomes large relative to m, the effectiveness of the special case decreases. Although the graphs show the effects for $k\leq 15$, the range of interest is for $k\leq 5$.



a. Relative Effectiveness of Split Outcome Special Case for Node A

b. Relative Effectiveness of Split Outcome Special Case for Node B

Figure 4. Split Outcome Versus General Case Data Requirements When Splitting an Outcome of A Into k Outcomes

Just as the ignored outcome special case, the split outcome special case shows exponential growth relative to the number of conditional predecessors. For both of

these special cases, however, the required number of probability assessments is some fraction of those required in the general case.

Changes in the Number of Variables. While the previous two special cases dealt with the expansion of the outcome space for a particular variable, now we examine the addition of a new variable into the diagram. While adding a variable to the diagram increases the number of probabilities in the joint distribution by a factor equal to the number of outcomes for the new variable, the increased number of additional probability assessments required for an influence diagram depends primarily on the conditioning changes brought about by the new variable.

For example, consider the following change to the metastatic cancer influence diagram shown previously in Figure 1. A new disease, denoted by the variable X with outcomes {present, absent}, is discovered which can cause brain tumors. The joint distribution, which had 2^5 probabilities under the original state of information, now contains 2^6 probabilities (an increase of 32). The effect on the marginal and conditional distributions is less, however. One probability must be assessed for the new node, and four $[(2-1)\times 2^2]$ probabilities must be assessed for node BT, increasing the total number of probabilities needed for the influence diagram from eleven (under the old state of information) to fourteen.

As indicated in this example, if the variable A is added to an influence diagram, the number of probability assessments required to define A's distribution depends on the number of outcomes for A and the number of outcomes for each conditional predecessor of A. Since there was previously no information in the diagram regarding A, all of these probabilities must be assessed. Additionally, the variables which now have A as a conditional predecessor (i.e., the direct successors of A) must now have their distri-

butions reassessed. One special case, the "assumed constant outcome" case, was identified that would reduce the number of required assessments.

Assumed Constant Outcome Special Case. One way that the number of assessments can be reduced is if the old state of information, &, is just the new state of information with the added condition that $A = a_0$. This might be the case when an expert learns that a factor previously considered constant did, in fact, have additional outcomes. Part of the probabilities, for nodes which gain A as a conditional predecessor, would then transfer directly from the original state of information to the new state of information. Suppose B is a node with p outcomes that is conditioned (under the new state of information, &') on the newly added A. If this special case applies, part of the new probability distribution for B is given by

$$P[B = b_j \mid A = a_0, C(B), \&'] = P[B = b_j \mid C(B), \&] \quad j = 1, \ldots, p \tag{14}$$

If there were only two outcomes of A, then the conditional distributions of all the direct successors of A would be half complete—a reduction of 50% in the number of assessments.

This special case provides no reduction in the number of assessments required to determine A's distribution, because originally there was no information about A in the influence diagram. If A has k outcomes, and each of A's conditional predecessors has n outcomes, then the number of assessments for both the general case and the constant outcome special case is $(k-1) \times n^{\|C(A)\|}$. If the assumed constant outcome special case is applicable to a direct successor like B, above, then the number of assessments required to determine the probability distribution of B decreases. If each conditional predecessor (other than A) of B has n outcomes, the number of assessments drops from $k \times (p-1) \times n^{\|C(B) \setminus A\|}$ for the general case to $(k-1) \times (p-1) \times n^{\|C(B) \setminus A\|}$ for the special case.

As seen for the other special cases, the data requirements grow exponentially as the number of conditional predecessors increases, but the original probabilities do not need to be reassessed when the assumed constant outcome case holds.

In practice, this special case is, perhaps, the easiest to see being applicable. Recall the KBS described earlier this chapter, which could be used to monitor intelligence reports and draw conclusions about possible enemy maneuvers and tactics. Suppose the original KBS was based on a state of information where the enemy clearly had the upper hand: better technology, a greater number of weapon systems, and better training, for instance. It is not at all unlikely that some, or all, of the probabilities of these maneuvers would change if the enemy were fighting in a situation where 1) they had the upper hand; 2) their opponent had the upper hand; or 3) neither side had the upper hand (equivalent capabilities). Depending on the capabilities of the opposing force, certain discriminators may indicate different tactics or maneuvers.

The graphs in Figures 5.a and 5.b show the relative number of assessments required for the assumed constant outcome special case. As in the previous graphs, these graphs show the ratio of the number of assessments required for the assumed constant outcome special case to the number required in the general case. Since this special case does not reduce the required number of assessments to determine the probability distribution of the newly added variable, Figure 5.a shows that exactly the same number of probabilities must be assessed (ratio = 1). Figure 5.b, like Figures 3 and 4, shows the decreasing effectiveness of this special case as k becomes larger for any nodes which gain A as a conditional predecessor and for which the special case applies. The ratio for this graph is given by $\frac{k-1}{k}$.

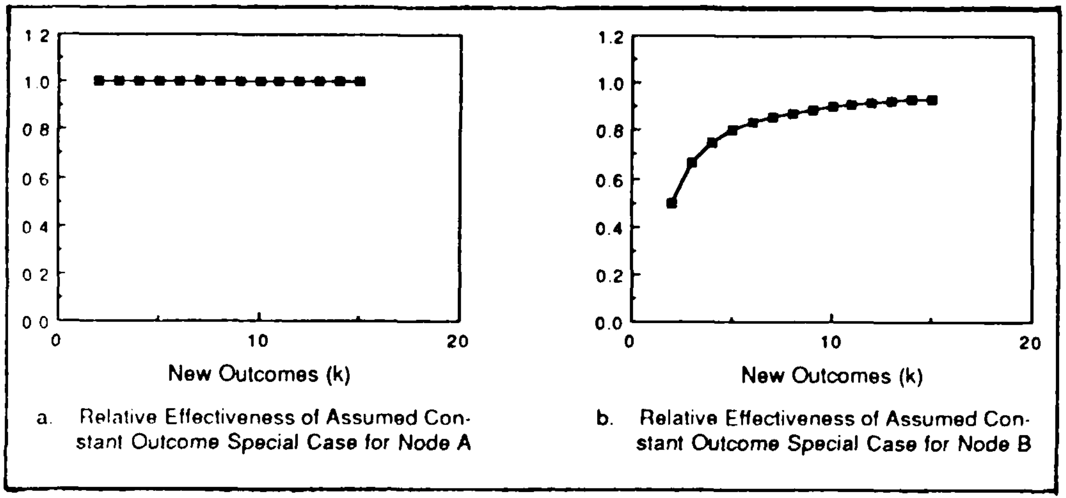


Figure 5. Assumed Constant Outcome Versus General Case Data Requirements When Adding New Variable A

Changes in Conditioning. When an arc is added between two nodes, say from A to B, only B must have its probability distribution reassessed. Since the distribution for the predecessor node (A) is defined by the conditional predecessors of A, no changes must be made to A's distribution. The effect on B, however, is one seen earlier, in the discussion concerning changes in the number of variables. In fact, adding an arc can be viewed as a special case of adding a new variable. Thus the assumed constant outcome special case may also be applicable when adding a new arc between existing nodes. For B, the new direct successor of A, the situation is identical to that for a successor of a newly added node: if the original probability distribution of B is valid for one outcome of A, those values need not be gathered again.

The Importance of Conditional Independence. The importance of getting the correct conditioning relationships in the influence diagram can not be overstated. If valid relationships are left out, the domain-specific knowledge base will be incomplete,

and the system may reach conclusions that differ greatly from the expert's. This would eventually lead to a maintenance action to correct the discrepancy, much as rule-based systems are updated when they reach incorrect conclusions. If unnecessary arcs are included in the diagram, the number of probabilities which must be assessed is unduly increased. The conclusions will be the same as those reached using the diagram without the unnecessary arc, but more steps (i.e., more computer resources) will be required to reach those conclusions.

Summary

Three special cases (ignored outcome, split outcome, and assumed constant outcome) point to situations where part or all of the marginal and conditional probabilities for nodes with changed incoming arcs can be used under the new state of information. Although these special cases do provide some decrease in the number of probability assessments which must be done to complete the modified influence diagram, their applicability must be determined, by the expert, for each change that is made. Since the number of required assessments, even when a special case applies, is exponential in the number of conditional predecessors to the node being reassessed, the selection of the minimum essential conditioning relationships is much more important in keeping the number of assessments as low as possible.

IV. Conclusions and Recommendations

Throughout the first three chapters, various aspects of probabilistic knowledge-based systems have been discussed, beginning with a discussion of the concerns which have hampered the use of probability in many KBS's. Existing efforts to alleviate those concerns, along with some of the advantages of using probabilities, were presented in Chapters I and II. Chapter III explored ways to reduce the amount of new data encoding when modifying a probabilistic knowledge-based system. This analysis was motivated by the perceived difficulties in the encoding of probabilistic data, and the difficulty of modifying a probabilistic KBS due to complex interactions among the variables. The following conclusions are drawn from this research.

Conclusions

Importance of Graphical Representations. The first conclusion, evident from discussions in Chapters II and III, is that the method chosen for representing the probabilistic information is extremely important, both in the use and the maintenance of the KBS. Recent research in this area has focused on the use of directed, acyclical graphs, which provide the following advantages:

1) Local modularity, provided by conditional independence within the graph, limits the number of reassessments which must be done: only nodes which have experienced a change to their outcome spaces or incoming arcs (conditional dependencies) must be reassessed. This reduces the growth from being exponential in the total number of variables in the graph to being exponential in the number of conditional predecessors for each node. This reduction could be quite significant, even for small influence diagrams. For instance, when adding an arc from the IC node to the H node in the metastatic cancer example, thirteen assessments are required to define all of the prior and conditional distributions in the diagram, but only four new assessments are needed to define H's new distribution.
2) Probabilities can be encoded and used in either the causal or evidential direction, as opposed the evidential direction required for rule-based systems.

3) Dependencies among variables can be represented easily, and more efficiently than rule-based representations.
4) Shachter's algorithm provides a simple method for finding any probability distributions of interest, conditioned on any set of evidences.

These advantages eliminate, or at least diminish, many of the concerns about using probabilities in KBS's.

Rule-Based Versus Graphical Representations. Given the advantages of graphical representations, one may think that such a representation would always be superior to a rule-based representation. This is not necessarily the case, especially when the dependencies found in the problem domain are not complex and the problem domain itself is relatively stable (not dynamic). As the problem domain becomes more dynamic, however, the ability to assess the probabilities in the direction (causal or evidential) most convenient to the expert makes the use of influence diagrams (or Bayesian networks) more desirable. Whether the problem domain is dynamic or not, as the dependencies in the knowledge base become more complex, the graphical representations again become more desirable.

Applicability of the Special Cases. Some reduction in the required number of probability assessments may be gained in situations where the special cases (ignored outcome, split outcome, or assumed constant outcome), discussed in Chapter III, apply. How often they will be applicable is unclear and will remain so until probabilistic KBS's are built, operated, and maintained over a period of time. However, the underlying assumptions for each special case seem to be reasonably realistic, so it would not be surprising to find that they apply in real decision problems.

Two characteristics of the special cases may limit their overall usefulness. First, the expert must determine their applicability to each change made in the knowledge

base: they may not (and probably will not) apply in many situations. Second, since the quantity of data contained in the original graph becomes smaller (relative to the total amount in the modified graph) as k increases, the effectiveness of the special cases decreases, requiring almost as many assessments as the general case. Given the advantages already provided by the graphical representations, the limited applicability and effectiveness of the special cases indicates that research aimed at using the original probabilities is less likely to yield significant results than research in other areas.

Areas for Future Research

Efficient Propagation Techniques for Multiply-Connected Graphs. One area which holds significant promise is the efficient and rapid propagation of probabilities throughout a multiply connected graph. As indicated in Chapter II, it is possible to apply Pearl's local propagation technique. but current methods suffer from exponential growth and an inability to provide a realistic interpretation for "auxiliary variables" added to transform the graph into a singly connected one. The question of interpretation may become a moot point if an algorithm can be developed which can convert a fully specified, multiply connected graph into a singly connected graph by adding auxiliary variables only to the internal portion of the graph, where they are neither seen as evidences nor of interest as hypotheses.

The Effect of Different Uncertainty Representations. Another area requiring much more research, as indicated by Henrion, is the effect of using different methods (probability, certainty factors, fuzzy set theory, etc.) to represent uncertainty in KBS's. Is there really any difference between conclusions reached using normative representations (i.e., probability) and those reached using descriptive representations, such as fuzzy set theory? If so, under what conditions are the differences pronounced, and which method

gives the "best" answer? If not, which is the easiest method to implement, or the most efficient?

More Efficient Probability Encoding Methods. For the maintenance of probabilistic knowledge-based systems, exploring ways to make the encoding of probabilities more efficient may be the most lucrative area to examine. Results from such a study would be equally applicable to artificial intelligence and decision analysis. Using a knowledge-based system to help guide, or to some extent automate, the encoding process is one possible approach.

Identification of Conditional Independencies. As indicated in Chapter III, unnecessary dependencies within an influence diagram increase the amount of effort, both in the building (probability assessments) and in the operation (mathematical computations) of a probabilistic KBS. While it is not likely that such unnecessary dependencies would be included in the original system, during maintenance efforts the expert and knowledge engineer may be taking a more restricted view of the system, focusing on a few particular nodes. This may prevent them from noticing that some of the arcs they are adding are unnecessary, given the other dependencies already in the system.

How can such unnecessary arcs be identified? The definition of conditional independence may provide the answer. If an arc from A to B is unnecessary, then the probabilities for outcomes of B are given by

$$P[B=b_j \mid A=a_i, C(B)\backslash A, \&] = P[B=b_j \mid C(B)\backslash A, \&] \tag{15}$$

for all outcomes a_i of A, and b_j of B. Using the strict equality may be unreasonable in actual use, however. It may be more reasonable to ask the expert to determine if an arc is necessary when this equation is approximately true. Checking for this approximate equality only needs to be performed when the probability distribution of a node is

either initially assessed, or modified during diagram maintenance. Such a routine can easily be added to systems like ALTERID, as a post-processor for the routine which accepts the node distributions from the knowledge engineer or the expert. Further research examining ways to bring these previously unrecognized conditional independencies, and other unrecognized implications of the probabilistic knowledge base, to the expert's attention may be beneficial. Recognizing these implications may give the expert a better understanding of the dependency structure, and reduce the number of required assessments for future diagram maintenance. Also, the impact of incorrectly asserting such underlying implications needs to be examined.

Summary

Attempts to keep the original probabilities usable, in some form, appear to be of limited value. There are, however, many other areas where research may provide interesting and possibly significant results. Research in these areas should enhance the ability of knowledge-based systems to reason under uncertain conditions, using normative decision models as the basis for recommended actions, and thereby increasing the usefulness of KBS in an uncertain world.

Bibliography

1. Breese, John S. *Knowledge Representation and Inference in Intelligent Decision Systems*. Ph.D. Dissertation. Stanford University, Stanford CA, 1987.

2. Heckerman, David E. and Eric J. Horvitz. "On the Expressiveness of Rule-based Systems for Reasoning with Uncertainty," *Proceedings AAAI-87: Sixth National Conference on Artificial Intelligence*. 121-126. Los Altos CA: Morgan Kaufmann Publishers, Inc. 1987.

3. Henrion, Max. "Uncertainty in Artificial Intelligence: Is Probability Epistemologically and Heuristically Adequate?" To Appear in "Expert Systems and Expert Judgment," *Proceedings of the NATO Advanced Research Workshop*. Porto, Portugal, August, 1986.

4. Howard, Ronald A. and James E. Matheson. "Influence Diagrams," *The Principals and Applications of Decision Analysis,* edited by R. Howard and J. Matheson. Menlo Park CA: Strategic Decisions Group, 1984.

5. Lindley, Dennis V. "The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence," *The Calculus of Uncertainty in Artificial Intelligence and Expert Systems: Proceedings of a Conference held December 28-29, 1984*. 1 December 1984—30 November 1985. 112-138. Grant N00014-85-G-0162. Washington DC: The George Washington University Institute for Reliability and Risk Analysis, 15 January 1986 (AD-A163 642).

6. ----. "Scoring Rules and the Inevitability of Probability," *International Statistical Review, 50:* 1-26 (1982).

7. Pearl, Judea. "Fusion, Propagation, and Structuring in Belief Networks," *Artificial Intelligence, 29:* 241-288 (September 1986).

8. ----. "Evidential Reasoning Using Stochastic Simulation of Causal Models," *Artificial Intelligence, 32:* 245-257 (1987).

9. Rich, Elaine. *Artificial Intelligence*. New York: McGraw-Hill Book Company, 1983.

10. Shachter, Ross D. "Probabilistic Inference and Influence Diagrams," Submitted to Operations Research (April 1986). Department of Engineering-Economic Systems, Stanford University, Stanford CA, Revised April 1987.

11. ----, and David E. Heckerman. "Thinking Backward for Knowledge Acquisition," *AI Magazine:* 55-61 (Fall 1987).

12. Shafer, Glenn. "Probability Judgement in Artificial Intelligence," *The Calculus of Uncertainty in Artificial Intelligence and Expert Systems: Proceedings of a Conference held December 28-29, 1984*. 1 December 1984—30 November 1985. Grant N00014-85-G-0162. Washington DC: The George Washington University Institute for Reliability and Risk Analysis, 15 January 1986 (AD-A163 642).

13. Waterman, Donald A. *A Guide to Expert Systems*. Reading MA: Addison-Wesley Publishing Company, 1986.

14. Zadeh, Lotfi A. "Fuzzy Sets and Possibility Theory," *The Calculus of Uncertainty in Artificial Intelligence and Expert Systems: Proceedings of a Conference held December 28-29, 1984*, 1 December 1984–30 November 1985. 78-98. Grant N00014-85-G-0162. Washington DC: The George Washington University Institute for Reliability and Risk Analysis, 15 January 1986 (AD-A163 642).

Vita

Captain Thomas F. Reid was born in Marlow, Oklahoma, 22 June 1960, to Aubra and Kathy Reid. He graduated from Marlow High School in 1978, and attended the University of Oklahoma, receiving the degree of Bachelor of Science in Mathematics in May 1982. He was commissioned upon graduation through the USAF ROTC program, and reported for active duty at Offutt AFB, Nebraska, on 20 October 1982 . There he served as Division Scientific Analyst for the Contingency Division of the Strategic Targeting Intelligence Center, 544th Strategic Intelligence Wing, until entering the School of Engineering, Air Force Institute of Technology, in May 1986.

Permanent Address: 412 N. Ninth
Marlow, Oklahoma 73055

END
DATE
FILMED
APRIL
1988
DTIC